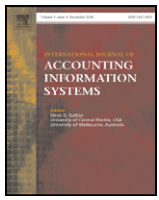
International Journal of Accounting Information Systems 12 (2011) 259–279

Contents lists available at SciVerse ScienceDirect

International Journal of Accounting Information Systems



The judgmental effects of strategy maps in balanced scorecard performance evaluations ☆

Rajiv D. Banker [a,1], Hsihui Chang [b,2], Mina Pizzini [c,*]

[a] Alter Hall 461, Fox School of Business and Management, Temple University, Philadelphia, PA 19122, United States
[b] 3600 Market St. 7th Floor, LeBow College of Business, Drexel University, Philadelphia, PA 19104-2875, United States
[c] 6212 Bishop Blvd. Cox School of Business, Southern Methodist University, Dallas, TX 75275-0333, United States

ARTICLE INFO

Article history:
Received 30 August 2010
Received in revised form 28 July 2011
Accepted 4 August 2011

Keywords:
Strategy maps
Balanced scorecard
Causal models
Performance evaluation

ABSTRACT

We examine whether supplemental information displays affect decisions made using a common strategic performance measurement system, the balanced scorecard. A distinguishing feature of the balanced scorecard (BSC) is the number and diversity of its metrics. To effectively formulate a decision from such a complex information set, managers must view these measures within their strategic context (Kaplan and Norton, 1993, 1996). However, academic studies indicate that problems in communication and comprehension of the strategic logic underlying the scorecard hinder its implementation and use (Lipe and Salterio, 2000; Malina and Selto, 2001; Ittner et al., 2003a, 2003b). We investigate whether a supplemental information display, in the form of a strategy map, results in performance evaluation judgments consistent with the recognition of relations between performance metrics and strategy. Strategy maps are causal diagrams depicting temporally-separate and non-linear relations between scorecard performance measures and overriding strategic objectives. As predicted, we find that performance evaluation decisions are more consistent with the achievement of strategic objectives when participants are provided with strategy maps.

1. Introduction

This study investigates whether supplemental information displays, in the form of strategy maps, affect decisions made using a common strategic performance measurement system, the balanced scorecard

☆ Data availability: Data and research instrument are available from the authors upon request.

* Corresponding author. Tel.: +1 214 768 3702; fax: +1 214 768 4099.
E-mail addresses: banker@temple.edu (R.D. Banker), hc336@drexel.edu (H. Chang), mpizzini@smu.edu (M. Pizzini).
[1] Tel.: +1 215 204 2029; fax: +1 215 204 5587.
[2] Tel.: +1 215 895 6979; fax: +1 215 895 6279.

1467-0895/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.accinf.2011.08.001

(BSC).[3] A strategic performance measurement system is a set of nonfinancial and financial objectives and performance measures representing a causal chain of activities that articulates management's hypothesis of strategy (Epstein and Manzoni, 1997). A distinguishing feature of the BSC is the number and diversity of its metrics: BSCs contain sixteen or more leading and lagging measures that capture performance along multiple dimensions (customer relations, internal processes, organizational learning and growth, and financial). The BSC's inherent complexity creates difficulties in communication and comprehension of its underlying logic which hinder implementation and use (Lipe and Salterio, 2000; Malina and Selto, 2001; Ittner et al., 2003a, 2003b; Ittner and Larcker, 2003; Banker et al., 2004; Dilla and Steingbart, 2005).

Research on causal modeling suggests that strategy maps can simplify and facilitate the transmission of complex systems, and thus implies that strategy maps have the potential to help decision makers overcome the cognitive challenges posed by the BSC (Fiol and Huff, 1992; Vera-Muñoz et al., 2007). This approach is consistent with studies in accounting, information systems, and psychology that find that altering or supplementing the manner in which information is presented can improve performance on decisions requiring complex judgments (e.g., Vessey, 1991; Tuttle and Kershaw, 1998; Lipe and Salterio, 2002; Dilla and Steingbart, 2005). Accordingly, the purpose of this study is to investigate the impact of supplemental strategy maps on BSC decisions using an experiment.

Strategy maps are causal maps depicting relations between BSC performance measures and overriding strategic objectives. Strategy maps can aid managerial decisions if they enable managers to assess a measure's relative importance to the achievement of strategic goals (*i.e.*, linkage to strategy) and thus provide cues for managers to weight and aggregate BSC measures in formulating an overall decision. Kaplan and Norton (2000, 2004, 2006) instruct managers to use strategy maps to communicate the BSC; and, several companies use strategy maps with the BSC (Mair, 2002; Kaplan and Norton, 2004, 2006; Urrutia and Eriksen, 2005; Veth, 2006). Commercial scorecard applications, such as Oracle Hyperion Performance Scorecard, even produce strategy maps.[4] Research in information systems and management has long recognized the value of causal maps in structuring complex problems (e.g., Axelrod, 1976; Eden et al., 1992; Fiol and Huff, 1992). Causal maps help individuals to construct more accurate mental models of complex systems (Johnson-Laird, 1983), and experimental research suggests that decision makers perform better when their mental models are more similar to the external systems they represent (e.g., Wyman and Randel, 1998; Davis and Yi, 2004; Capelo and Dias, 2009).

Our experiment, which is based upon Banker et al. (2004), requires participants to use the BSC to evaluate managers in two different business units of the same company. Each business unit has its own 16-measure BSC. The measures differ regarding their relationship to business unit strategy: some are causally linked to strategy (linked) and some are not (non-linked). Banker et al. (2004) investigated whether decision makers relied more on strategically linked measures than non-linked measures or measures that are common across scorecards.[5] They manipulated detailed strategic information at two levels: (1) no strategic information (Benchmark Treatment), and (2) a narrative on strategy and a strategy map (Strategy Information Treatment). Our research question and experimental design differ from Banker et al. (2004) in that we investigate whether supplementing narrative strategic information with a strategy map leads to decisions that are more consistent with strategic objectives. Thus, we add a third treatment in addition to the two treatments used in Banker et al. (2004). Participants in this third treatment receive narrative strategy information, but not a strategy map (Narrative Treatment). This enables us to investigate whether a strategy map, which is informationally equivalent to the narrative regarding unique BSC measures, improves decision-making. Consistent with our predictions, we find that participants that receive both a strategy map and a narrative on strategy rely more on strategically linked measures than those that receive only a narrative. This result suggests that strategy maps help managers view performance measures within their strategic context, and thereby, reduce the cognitive difficulty of using the BSC for performance evaluation.

[3] Kaplan and Norton (1992) developed the Balance Scorecard in the early 1990s. Management Tools & Trends survey found that over 55% of companies surveyed in the US and Western Europe use a balanced scorecard (www.bain.com/management_tools/).

[4] http://www.oracle.com/appserver/business-intelligence/hyperion-financial-performance-management/hyperion-performance-scorecard.html.

[5] Lipe and Salterio (2000) found that experiment participants using the BSC based evaluations on performance measures that were common across different business units and ignored measures that were unique to different individual business units when they were not given business-unit specific strategy information. Banker et al. (2004) found that supplying participants with strategy information specific to each business unit helped participants overcome common-measure bias.

This research makes three contributions to the literature. First, it extends prior studies that investigate methods to improve managers' ability to effectively use the BSC framework to make decisions (Banker et al., 2004; Libby et al., 2004; Webb, 2004; Dilla and Steingbart, 2005). In particular, our findings suggest that Banker et al. (2004) would not have obtained the results they did without the inclusion of a strategy map. Accordingly, this study is of interest to the many managers who use scorecards and other cognitively demanding strategic performance measurement systems. Second, our results indicate that a causal map facilitates the transmission of complex ideas from individual to individual, and hence support a map's decision usefulness for individuals *other than those who formulate it*. Thus our results support causal mapping as a communication tool, in addition to a problem-structuring tool. Finally, our results are consistent with studies in the information systems literature that find decision makers must understand an information system to fully reap its benefits (e.g., Davis, 1989; Davis et al., 1989; Subramanian, 1994; Venkatesh and Davis, 2000).

The next section reviews related literature and presents the research hypothesis. We describe our experimental design in Section 3 and present results in Section 4. Conclusions are discussed in Section 5.

2. Literature review and hypothesis

2.1. Using the balanced scorecard for performance measurement

Kaplan and Norton (1992) developed the balanced scorecard (BSC) to supplement traditional financial measures with operating performance measures oriented toward customers, internal processes, and learning and growth activities. Balanced scorecards typically contain between 16 and 28 different measures grouped into four-to-six categories. The measures are both non-financial (e.g., cycle time, customer satisfaction) and financial (e.g., sales growth, debt ratios) and include both leading indicators (e.g., new product introductions) and lagging indicators (e.g. profit margin) of performance. Each strategic business unit (SBU) within a firm has its own scorecard reflecting its distinct strategy. Kaplan and Norton (1993, 1996) contend that scorecards improve performance by translating strategy into tangible objectives that are linked in a causal chain of leading and lagging indicators covering the different scorecard perspectives. This is consistent with claims that the articulation of such linkages is critical because intangible assets have become increasingly important to a firm's success (e.g., Eccles, 1991; Brancato, 1995; Fisher, 1995; Young and O'Byrne, 2001).

Field research suggests that the BSC should be incorporated into compensation and performance evaluation (Rucci et al., 1998; AICPA and Maisel, 2001; Kaplan and Norton, 2001; Malina and Selto, 2001). However, Kaplan and Norton (1996) do not specify exactly how evaluators should combine the many disparate performance measures to formulate \an overall performance assessment. In fact, they advocate subjective assessment in performance evaluation, contending that a subjective reward system is "easier and more defensible to administer and also less susceptible to game playing" (Kaplan and Norton, 1996, 220). Hence, evaluators are left to use their own judgment in assigning weights to various measures. Although the weights should be driven by the importance of each measure with regard to each SBU's strategy, extant research suggests that managers fail to appreciate the cause–effect logic built into the BSC. For example, Lipe and Salterio (2000) reported that experiment participants using scorecards to evaluate the managers of two different SBUs did not consider the unique strategy of each SBU and simply relied on measures common to each SBU. Ittner et al. (2003a) found that managers systematically overweighed financial measures relative to non-financial measures because managers did not understand strategic goals and their connection to managers' actions. Ittner and Larcker (2003) assert that firms commonly fail to reap the rewards of performance measurement systems that include non-financial measures, such as the BSC, because they do not explicitly identify causal relations between managers' immediate actions and goals and firm-wide strategic objectives.

The number of problems encountered using the BSC is not surprising given its complexity. Relative performance evaluation in the BSC framework requires decision makers to evaluate multiple performance measures (16–28) for two or more strategic business units. To subjectively formulate a solution, decision makers must understand relationships between performance measures and their relevance to the overall strategy. Recognition of relationships between temporally-separated (*i.e.*, leading and lagging) performance indicators and non-linear relationships arising from non-financial measures places

extra cognitive demands on memory and attention (Einhorn and Hogarth, 1986; Sterman, 2000). Given the number of measures and the complex nature of their relationships, a human decision maker cannot make a decision analytically (Umanath and Vessey, 1995). Since there is no objective, computational procedure available, decision makers must spend time thinking about what to do and searching for solutions (Benbasit and Lim, 2000). Graphical representation can facilitate problem comprehension by summarizing information and providing an overall perspective in complex, less analyzable tasks (Desanctis and Jarvenpaa, 1989).

Dilla and Steingbart (2005) designed an experiment to examine whether tabular and graphical displays could help managers overcome the difficulties associated with using the BSC for performance evaluation. Drawing on cognitive fit theory, they hypothesized that supplementary tabular and graphical displays would improve judgment consensus and consistency. They found that supplemental tabular displays improved consistency, but not consensus, while supplemental graphical displays decreased judgment consensus, but did not affect consistency. They concluded that supplemental display format is important and that care must be taken when designing and implementing BSC decision aids. We build upon Dilla and Steingbart (2005) by investigating supplemental displays in the form of strategy maps.

2.2. Causal maps and mental models

A strategy map for a BSC is a causal map depicting relations between various performance measures and corporate objectives. Causal maps express the judgment that certain events or actions will lead to particular outcomes. Research in cognition has long recognized the value of causal maps in structuring complex problems (e.g., Axelrod, 1976; Eden, 1992; Fiol and Huff, 1992), and several studies have addressed the construction of causal maps (e.g., Eden et al., 1992; Tan and Hunter, 2002; Scavarda et al., 2006; Druckenmiller and Acar, 2009). Fewer studies have investigated the impact of causal maps on complex strategic tasks. Such research has generally used a case-study approach to examine the decision-making performance of the individual(s) that formulated the map, and thus reflects performance effects obtained through the non-trivial exercise of developing the map (e.g., Wyman and Randel, 1998; Scavarda et al., 2006; Barad and Dror, 2008; Montibeller et al., 2008; Capelo and Dias, 2009). However, a claimed strength of the map is its ability to both simplify and facilitate the transmission of complex ideas from individual to individual (Fiol and Huff, 1992). Accordingly, we investigate the decision-relevance of maps for individuals who did not formulate the map.

Causal maps have the potential to improve decision performance by helping individuals to construct more accurate mental models of multi-attribute systems. Mental models are simplified cognitive representations of the structure of an external system and are constructed based on an individual's understanding of how pieces of information interrelate (Johnson-Laird, 1983). Experimental research suggests that decision makers perform better when their mental models are more similar to the external systems they represent (e.g., Wyman and Randel, 1998; Davis and Yi, 2004; Capelo and Dias, 2009). The ability to form an accurate mental model reflects knowledge organization rather than superior perceptual ability (Wyman and Randel, 1998). Thus, even when individuals have equal knowledge, differences in knowledge organization (mental models) potentially explain differences in performance.

Different ways of presenting the same information can lead to differences in knowledge organization and different mental models in memory. Causal maps permit a rich representation of ideas through the modeling of complex chains of arguments (Fiol and Huff, 1992; Montibeller and Belton, 2006; Montibeller et al., 2008). Causal-model theory contends that people preferentially learn in the cause–effect direction (Waldmann et al., 1995), thus directional arrows should help decision makers build more accurate and comprehensive mental models of complex systems. Directional arrows also facilitate the cognitively demanding task of identifying temporally-separated (leading and lagging) and non-linear cause–effect relationships. Focusing on cause–effect relations between specific measures reduces the cognitive difficulty required to understand the complete covariation matrix among all BSC measures (Waldmann et al., 1995). Vera-Muñoz et al. (2007) found that accountants prompted with causal-model information were better able to interpret and exploit cause-and-effect covariation information implied by benchmark data. Fiol and Huff (1992) contend that maps focus attention and highlight priorities, which are critical when managers face too much information. In this way, strategy maps can help reduce the cognitive load imposed by the BSC.

2.3. Hypothesis

To effectively use the BSC, decision makers must assess each measure's relative importance to the achievement of strategic goals. As discussed previously, research in information systems, psychology, and management suggests that strategy maps provide appropriate cues for managers to combine and weight various BSC measures in formulating a decision. Accordingly, managers who have strategy maps will perceive strategically linked performance measures to be more useful for performance evaluation than managers who have only narrative strategy information and those that have no detailed strategy information. Furthermore, participants that have both narrative and graphical strategy information will also recognize that non-linked measures are of little value and thus will place less weight on them in evaluating performance than participants who have no information or only narrative information. This leads to the following hypothesis:

H. When evaluating performance, decision makers who have narrative strategy information and strategy maps will place greater (less) weight on strategically linked (non-linked) measures than those who have only narrative strategy information or no strategy information.

3. Experimental design

3.1. Overview of experiment

The structure of the experiment follows Lipe and Salterio (2000), which asked participants to evaluate the performance of two divisional managers using a BSC. Experimental materials come from Banker et al. (2004), which investigated the effect of strategically linked and non-linked measures on managers' performance evaluation judgments. Participants in the experiment read a case asking them to assume the role of a senior executive at Smithson Stores, a clothing retailer that recently implemented a performance measurement system based upon the balanced scorecard. The case discusses two strategic business units (SBUs) of Smithson Stores, The Women's Store and The Family Store, each with a distinct strategy and target market. The case provides a narrative overview of the company, which includes Smithson's mission statement and a brief description of each SBU (Exhibit 1). The Women's Store is described as "a retailer of fashionable women's clothing … in the mature stage of the business lifecycle," and The Family Store is described as "a retailer catering to young families … in the growth stage of the lifecycle." The case also includes a balanced scorecard for each SBU that shows actual versus targeted performance for 16 performance measures (Exhibit 2).

Participants were required to evaluate the performance of both managers based upon the information provided in the case. Participants ranked each manager on a 13-point scale anchored as follows:

0 = Reassign: sufficient improvement unlikely;
2 = Very Poor: considerably below expectations;
4 = Poor: somewhat below expectations;
6 = Average: meets expectations;
8 = Good: somewhat above expectations;
10 = Very Good: considerably above expectations; and,
12 = Excellent: far beyond expectations, manager excels.

Participants were then asked to recommend one of the two managers for promotion to the position of Vice President. Finally, participants provided demographic information and answered questions regarding strategy comprehension, case understanding, and task difficulty.

3.2. Experiment participants

One-hundred-and-eighty students enrolled in the MBA core course in accounting completed the experiment in the one hour allotted for it. Fifty-four percent of students were male, 69% had lived in the U.S. for at least five years, and 96% had visited a retail clothing store in the past 12 months. Students' mean age was 30 years old, they had an average of seven years of full-time work experience, and 17% had retail experience.

Exhibit 1
Narrative Overview of Company
(provided to all participants)

Smithson Stores

Smithson Stores is a major U.S. clothing retailer operating 8 independent retail chains with over 3,000 stores and $6 billion in annual sales. Each chain is organized as a separate strategic business unit (SBU), with a distinct image and target market. Smithson's mission statement is presented below:

> *We will be an outstanding apparel supplier in each of the individual niche markets we serve.*

At Smithson's, corporate management makes all financing decisions and sets overall financial objectives, but grants SBU managers the freedom to tailor SBU strategy to their respective markets. This decentralized structure gives the chains the flexibility to respond to the unique trends and opportunities that arise in their individual markets. However, decentralization creates problems in evaluating SBU performance. Smithson's has traditionally used the same, financially-oriented performance measures across all of its SBUs, regardless of strategy or market. Unfortunately, a measure that was excellent for one SBU often provided the wrong incentives for another. To address this concern, the company instituted a new performance measurement system based upon the Balanced Scorecard. In a recent meeting, Smithson's CEO announced,

> *Our decentralized structure has been the bedrock of our continued marketing success. However, SBU managers must be able to relate SBU-level performance measures to our overall corporate goals of 25% annual sales growth and 50% return on sales. Thus, we will use the Balanced Scorecard to develop a set of SBU-level performance measures that provides coordination and direction to our diverse retail chains in reaching this goal, without sacrificing managers' independence.*

Accordingly, corporate and business-unit managers together developed separate Balanced Scorecards (BSCs) for each individual SBU. The scorecards for two of Smithson's SBUs, The Women's Store and The Family Store, are detailed in Exhibits 1 and 2, respectively. The Women's Store, a retailer of fashionable women's clothing, is in the "mature" stage of the business lifecycle. The Family Store, which caters to young families, is still in the "growth" stage of the lifecycle.

3.3. Design and procedure

The experiment employs a $2 \times 2 \times 3$ between subjects design, in conjunction with a two-level within-subjects factor (Exhibit 3). The first factor refers to an SBU's relative performance on measures that are directly linked to strategy (*i.e.*, linked). The Women's Store could perform better on strategically linked measures (Link-WS) or The Family Store could perform better on linked measures (Link-FS). Similarly, the second factor reflects relative performance on measures that are not directly linked to the SBU's individual strategy (*i.e.*, non-linked). The Women's Store could dominate The Family Store on non-linked measures (Non-WS); or, The Family Store could dominate on non-linked measures (Non-FS). The third factor refers to the information treatment to which each participant was assigned as follows: Benchmark, Narrative, or Strategy Map. Each participant evaluated both managers, producing a within-subjects factor.

3.4. Manipulation of BSC measures

We designed a separate 16-measure scorecard for each of Smithon's two SBUs (Exhibit 2). We specified four measures for each of the four BSC perspectives (financial, customer, operational, learning and

Exhibit 2
Balanced Scorecard for The Women's Store Used in the Experiment

Measure[1]	Target	Actual	Percent Better than Target
Financial:			
1. Sales margins	60%	67.02%	11.70%
2. Sales growth per store	15%	16.75%	11.67%
3. Inventory turnover	6	6.59	9.83%
4. Debt-to-assets ratio	< 20%	18.07%	9.65%
Customer:			
1. Price relative to competitors' price	+7%	7.79%	11.29%
2. Customer satisfaction rating	80%	88.44%	10.55%
3. Sales per square foot of retail space	$30,000	$33,090	10.30%
4. Number of credit card customers per store	8,000	8,911	11.39%
Internal Process:			
1. Brand recognition rating	80%	87.60%	9.50%
2. Number of stock-outs	< 3 times	2.66	11.33%
3. "Mystery Shopper" audit rating	85%	93.47%	9.96%
4. Time to process customer returns	< 4 min.	3.54	11.50%
Learning and Growth:			
1. Employee satisfaction	80%	87.96%	9.95%
2. Employee suggestions per year	2.5 times	2.74	9.60%
3. Store computerization	60%	66.24%	10.40%
4. Hours of training invested in brand managers each year	80 hours	89.10	11.38%

Balanced Scorecard for The Family Store Used in the Experiment

Measure[1]	Target	Actual	Percent Better than Target
Financial:			
1. Sales margins	45%	48.41%	7.58%
2. Percentage of sales from new stores	60%	64.56%	7.60%
3. Operating cash flow as a percentage of sales	5%	5.27	5.40%
4. Debt-to-assets ratio	< 50%	47.35%	5.30%
Customer:			
1. Price relative to competitors' price	+2%	2.15%	7.50%
2. Customer satisfaction rating	70%	74.31%	6.16%
3. Percentage of sales to new customers	5%	5.30%	6.00%
4. Number of new items in which first to market	10	10.75	7.50%
Internal Process:			
1. Market share per advertising dollar	5%	5.29%	5.80%
2. Vendor rating	75%	80.25%	7.00%
3. "Mystery Shopper" audit rating	85%	90.37%	6.32%
4. Time to process customer returns	< 4 min.	3.71	7.25%
Learning and Growth:			
1. Employee satisfaction	75%	79.17%	5.56%
2. Sales associates with college degrees	25%	26.40%	5.60%
3. Store computerization	80%	85.12%	6.40%
4. Retail experience of marketing managers	8 yrs.	8.56	7.00%

[1] Participants were provided with definitions of each performance measure.

Exhibit 3
Experimental Design: Manipulation of Balanced Scorecards and Assignment of Participants to Information Treatments

The experiment employs a 2 x 2 x 3 between subjects design. The first factor refers to an SBU's relative performance on measures that are directly linked to strategy (linked favor The Women's Store or The Family Store); and, the second factor reflects relative performance on non-linked measures (non-linked favor The Women's Store or The Family Store). Each participant was provided with a pair of BSCs (one for each SBU). Actual performance was varied within the BSC pairs to correspond to the four possible performance measure type-favored SBU combinations. Forty-five participants received Combination 1, another 45 received Combination 2, etc. Strategy information provided to the 45 participants in each combination grouping varied at 3 levels: 15 received no SBU-specific strategy information (Benchmark Treatment), 15 received only narrative information on strategy (Narrative Treatment), and 15 received both narrative information and a strategy map (Strategy Map Treatment).

		Non-linked measures favor[1]	
		The Women's Store	The Family Store
Linked measures favor[1]	The Women's Store	**Combination 1: Link – WS, Non – WS** 15 participants – Benchmark 15 participants – Narrative 15 participants – Strategy Map	**Combination 2: Link – WS, Non - FS** 15 students – Benchmark 15 students – Narrative 15 students - Strategy Map
	The Family Store	**Combination 3: Link – FS, Non - WS** 15 students - Benchmark 15 students - Narrative 15 students - Strategy Map	**Combination 4: Link – FS, Non - FS** 15 students - Benchmark 15 students - Narrative 15 students - Strategy Map

[1] Link – WS: linked measures favor The Women's Store.
Non – WS: non-linked measures favor The Women's Store.
Link – FS: linked measures favor The Family Store.
Non – FS: non-linked measures favor The Family Store.

growth). Within each perspective, two measures were directly linked to SBU strategy and two measures were not (Exhibit 4). We included performance measures that were not directly linked to strategy because companies often include general, readily-available metrics in evaluation even when those measures are not particularly relevant (Malina and Selto, 2001; Ittner and Larcker, 2003; Bryant et al., 2004; Burney and Widener, 2007). We compare measures that are clearly linked to strategy with those that are not to determine whether strategy maps enable decision makers to distinguish good performance that leads to strategic objectives from good performance that is not directly related to strategic goals.[6]

Since each participant evaluated both SBU managers, we provided each participant with a pair of scorecards, one for each SBU. We manipulated actual performance within each pair of scorecards to correspond to one of the four performance measure type-favored SBU combinations as follows (Exhibit 3):

Combination 1: Link-WS, Non-WS;

[6] Whether directly linked to the strategy or not, a better rating on any performance measures is a positive signal. For example, although inventory turnover is not directly linked to strategy for The Women's Store, ceteris paribus, higher turnover is better than lower turnover.

Exhibit 4
Linked and Non-linked Performance Measures for The Women's Store and The Family Store Balanced Scorecards

Type[1]	Measures
Financial Measures:	
Link	Sales margins
Non	Debt-to-assets ratio
Link, WS	Sales growth per store
Non, WS	Inventory turnover
Link, FS	Percentage of sales from new stores
Non, FS	Operating cash flow as a percentage of sales
Customer Measures:	
Link	Customer satisfaction rating
Non	Price relative to competitors' price
Link, WS	Sales per square foot of retail space
Non, WS	Number of credit card customers per store
Link, FS	Percentage of sales to new customers
Non, FS	Number of new items in which first to market
Internal Process Measures:	
Link	"Mystery Shopper" audit rating
Non	Time to process customer returns
Link, WS	Brand recognition rating
Non, WS	Number of stock-outs
Link, FS	Market share per advertising dollar
Non, FS	Vendor rating
Learning and Growth Measures:	
Link	Employee satisfaction
Non	Store computerization
Link, WS	Hours of training invested in brand managers each year
Non, WS	Employee suggestions per year
Link, FS	Retail experience of marketing managers
Non, FS	Sales associates with college degrees.

[1] Link indicates a linked measure for both The Women's and The Family Stores.
Non indicates a non-linked measure for both The Women's and The Family Stores.
Link, WS indicates a linked measure for The Women's Store.
Non, WS indicates a non-linked measure for The Women's Store.
Link, FS indicates a linked measure for The Family Store.
Non, FS indicates a non-linked measure for The Family Store.

Combination 2: Link-WS, Non-FS;
Combination 3: Link-FS, Non-WS; and,
Combination 4: Link-FS, Non-FS.

Following Lipe and Salterio (2000), we required that the eight performance measures of the same type (linked or non-linked) within a BSC favor the same SBU. We chose actual performance measure values such that each SBU performed better than its target for all measures on the scorecard; however, the amount by which actual SBU performance exceeded the target varied to accommodate the four performance measure type-favored SBU combinations. For example, Exhibit 2 contains a BSC pair reflecting Combination 1, in which The Women's Store outperforms The Family Store on linked and non-linked measures. Actual performance exceeds target on all measures in both scorecards; however, excess performance for The Women's Store

Exhibit 5
Strategy of Two Strategic Business Units of Smithson Stores
(provided to participants in the Narrative and Strategy Map Treatments)

Presented below is a brief description of the strategy of The Women's Store and The Family Store.

The Women's Store

The Women's Store (TWS) is an established specialty retailer that caters to fashion-conscious professional women. TWS's strong cash flows will be needed to fuel growth in Smithson's younger SBUs. TWS will contribute to corporate objectives by using its existing store network to further penetrate its target market, while improving margins and cash flows. Accordingly, sales growth will come through the introduction of new clothing lines, such as its business casual line, and excellent in-store shopping assistance designed to accommodate style-conscious, time-constrained customers. TWS will leverage its distinctive brand image to drive new clothing sales and margin growth within its target market, which is not very price sensitive. TWS's goal is to become a store in which women can shop for all of their wardrobe needs, from clothing to accessories to shoes, in one convenient location. By developing skilled brand managers who can broaden its product line to accommodate "one stop shopping" and by offering greater in-store shopping assistance, TWS hopes to compete with catalogue and on-line retailers. Central to TWS's growth strategy is the creation of a "perfect in-store shopping experience" that will entice busy women to visit the store, rather than shop on-line or via the phone.

The Family Store

The Family Store (TFS), which opened its doors three years ago, is one of several retail chains in the Smithson portfolio that is still in the "growth" stage of the business lifecycle. TFS carries classic, high-quality casual clothing, such as khakis, jeans, and polo shirts for men, women, and children. Although TFS offers stylish, high-quality merchandise, it is more of "fashion follower" than a "fashion leader". Therefore, unlike The Women's Store, which seeks to introduce many new product lines, TFS identifies a relatively narrow set of basic, functional styles each season and offers these items in a variety of colors and sizes. Management expects this new, high-end retailer to drive a significant portion of Smithson's overall growth, as TFS plans to double the number of stores over the next two years.

TFS's target customers are primarily young, upper middle-class families with significant disposable income, but little free time. TFS's combination of classic, high-quality men's, women's and children's apparel enables busy parents to shop for the entire family in one location. To attract this customer segment to its store, TFS is building an experienced marketing team that will launch an aggressive advertising campaign featuring memorable, humorous commercials appealing to all ages. Additionally, TFS plans to exceed customers' shopping expectations with its fun "in-store" atmosphere, visually appealing displays, and excellent service.

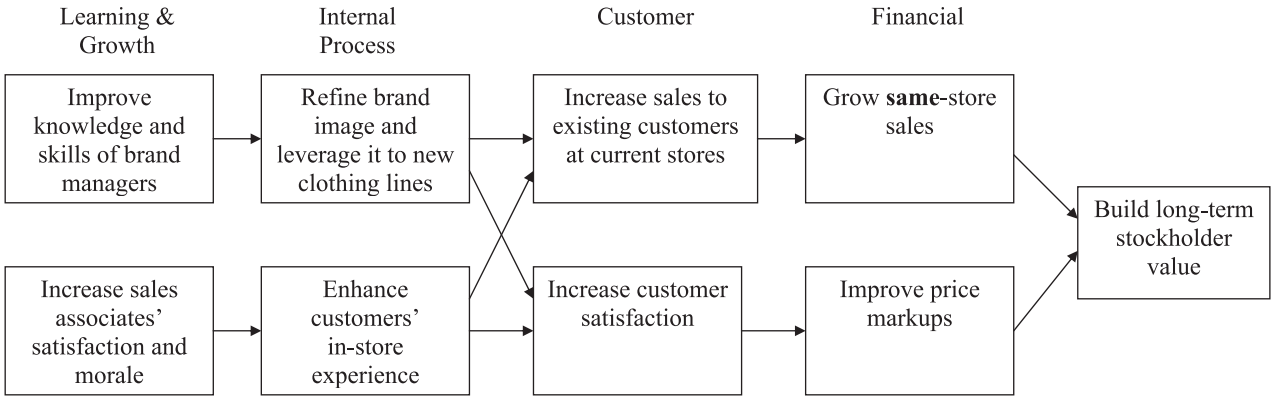
(percentage above target) is greater than excess performance for The Family Store on all measures. Each measure type had approximately the same excess performance. The average percentage above target was 10.6% for the better-performing SBU and 6.5% for the worse-performing SBU for both measure types.

3.5. Manipulation of information provided to participants

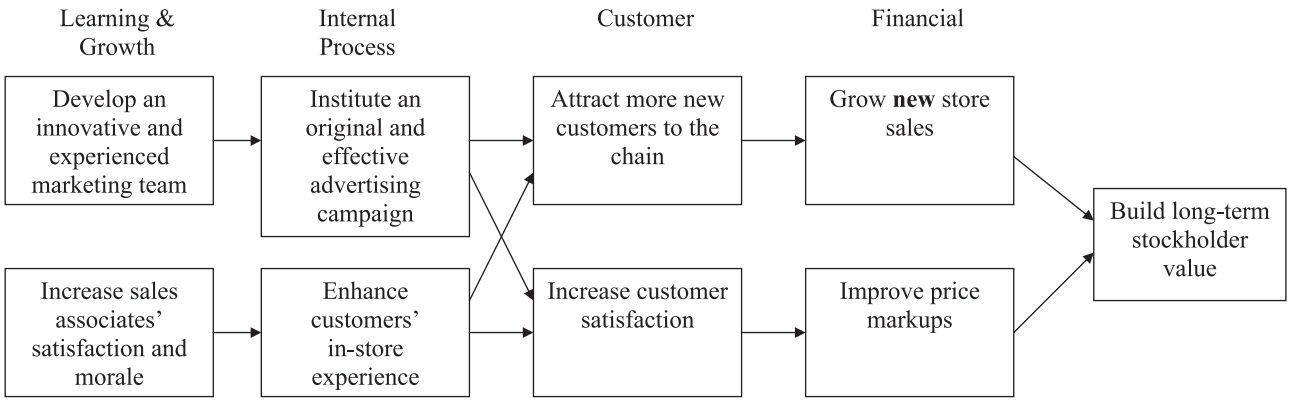
To assess whether information about business unit strategy affects performance evaluation, we created three information treatments: Benchmark, Narrative, and Strategy Map. Students assigned to the Benchmark Treatment received only the written overview of the company contained in Exhibit 1. Students assigned to the Narrative Treatment received the written overview from Exhibit 1, along with a brief

Exhibit 6 *(provided only to participants in strategy map treatment)*

The Women's Store: Strategic Objectives



The Family Store: Strategic Objectives



narrative outlining each individual SBU's strategy (Exhibit 5). Students assigned to the Strategy Map Treatment received the company overview (Exhibit 1), a brief narrative on each SBU's strategy (Exhibit 5), and a simple strategy map (Exhibit 6). The strategy map for each SBU was designed to be informationally equivalent to its corresponding strategy narrative in that all strategy map information that was unique to each SBU could be inferred from the narrative (Larkin and Simon, 1987). For example, The Women's Store narrative (Exhibit 5) states that this SBU will contribute to Smithson's overall growth by 'using its existing store network to further penetrate its target market" and "will leverage its distinct brand image to drive new clothing sales and margin growth within its target market." The SBU will achieve greater market penetration, "By developing skilled brand managers who can broaden its product line." These statements are captured in the upper portion of The Women's Store strategy map (Exhibit 6), and are reflected in the following linked measures: sales growth per store, sales per square foot of retail space, brand recognition rating, and annual hours of brand manager training.

To effectively use the BSC, managers must be able to evaluate each measure's relative importance to the achievement of an organization's strategic goals. Accordingly, managers must have a thorough understanding of strategy. In aggregating multiple BSC measures for performance evaluation, managers should place more weight on measures that have greater strategic significance. In the experiment, this corresponds to participants placing greater weight on linked measures when judging the performance of each SBU manager. We expect that the strategy map will facilitate strategy comprehension, and hence participants in the Strategy Map Treatment will be better able to recognize the strategic significance of linked measures than participants in the other treatments. Consequently, participants in the Strategy Map Treatment will rely more on strategically linked measures when evaluating performance. Therefore, comparisons across treatments are used to determine whether strategy maps affect performance evaluation.

It is important to note that we do not test whether a graphical format is superior to a narrative format (or *vice versa*). Rather, we test whether a supplemental graphical display, in the form of a strategy map, leads to more effective use of the BSC than a narrative alone does. The strategy map is intended to provide guidance in recognizing each performance metric's strategic significance, and thereby facilitate decision making with the BSC; the benchmark case serves as a control. Our approach follows Dilla and Steingbart (2005) and is similar to Frownfelter-Lohrke (1998), who compared judgment accuracy for graphical and tabular formats separately to a combination format that included both graphs and tables. This approach has also been used in the decision aid literature (e.g., Adams et al., 1992; Ashton, 1992; Todd and Benbasat, 1992; Eining et al., 1997).

3.6. Pilot tests

As previously discussed, we obtained the experimental materials from Banker et al. (2004), who pilot tested the materials with 45 Executive M.B.A. students. Participants rated their agreement (on an 11-point scale, where −5 indicates "strongly disagree" and +5 indicates "strongly agree") with statements assessing the usefulness of performance measure categorizations, the understandability of the case, and the realism of the case. Mean agreement ratings were significantly greater than zero for each question (all $p<0.001$, two-tailed), suggesting that the case was realistic and easy to understand. Further analysis of participant responses indicated no *a priori* difference between the perceived relevance of the set of strategically linked measures and the set of non-linked measures for each store.

3.7. Dependent measure

Each participant rated the performance of The Women's Store manager and The Family Store manager on a 13-point scale. The dependent variable in our hypothesis test is the difference between the evaluation scores the same participant assigned to each manager. Recall that we manipulated the pattern of performance for each BSC pair to correspond to the four performance measure type-favored SBU combinations (Link-WS, Non-WS; Link WS, Non-FS; Link-FS, Non-WS; Link-FS, Non-FS). Thus, the difference between a participant's ratings for each SBU manager captures the difference in weights the participant placed on linked and non-linked measures and reflects the extent to which the participant incorporated linked and non-linked measures into her evaluation. This difference is denoted $(W_i - F_i)_{COMBO,TREATMENT}$, where

Table 1
Mean ratings differences for each BSC combination and treatment group.

$$\Delta RATINGS_{COMBO,TREATMENT} = \left[\Sigma_i^{15}\left(W_{i,}-F_i\right)\right]/15,$$

where W_i and F_i are participant i's rating of The Women's Store and The Family Store managers, respectively; and, participant i received the BSC combination, COMBO, and TREATMENT indicated.

		Mean of ratings differences by *TREATMENT* (significance)[a]		
BSC combination (*COMBO*)	Performance measure type-favored SBU[b]	Benchmark treatment	Narrative treatment	Strategy map treatment
1	Link-WS, Non-WS	1.400 (0.001)***	1.333 (0.002)***	1.467 (0.001)***
2	Link-WS, Non-FS	0.267 (0.390)	0.667 (0.106)	0.867 (0.018)**
3	Link-FS, Non-WS	−0.133 (0.737)	−0.400 (0.320)	−0.933 (0.017)**
4	Link-FS, Non-FS	−1.200 (0.001)***	−0.867 (0.010)***	−1.200 (0.001)***

** and *** indicate significance levels of 5% and 1%, respectively.
[a] Link-WS: linked measures favor The Women's Store. Non-WS: non-linked measures favor The Women's Store. Link-FS: linked measures favor The Family Store. Non-FS: non-linked measures favor The Family Store.
[b] Significance levels refer to two-sided means tests for paired data to evaluate the null hypothesis that $\Delta\ RATINGS_{COMBO,TREATMENT} = 0$.

W_i and F_i are subject *i*'s rating for The Women's Store and The Family Store, respectively; and, subject *i* received the BSC combination (*COMBO*) and treatment (*TREATMENT*) indicated.

Table 1 contains the average rating differences given by the 15 participants in each of the 12 BSC combination-treatment groups, calculated as shown in Eq. (1):

$$\Delta - RATINGS_{COMBO,TREATMENT} = \left[\Sigma_i^{15}\left(W_{i,}-F_i\right)\right]/15 \quad (1)$$

For example, Δ-$RATINGS_{2,SM}$ is the average ratings difference reported by the 15 participants who received BSC Combination 2 and were in the Strategy Map Treatment. As shown in the second row of Table 1 (last column), the value of Δ-$RATINGS_{2,SM}$ is 0.867 ($p = 0.018$). This indicates that participants in the Strategy Map Treatment who received Combination 2 (Link-WS, Non-FS) assigned The Women's Store manager a rating that was on average 0.867 higher than the rating they assigned to the The Family Store manager. If Δ-$RATINGS_{COMBO,TREATMENT}$ is negative, ratings for The Family Store manager exceed those

Table 2
Manipulation check: mean responses to questions regarding participants' understanding of task by treatment.[a]

				Difference in treatments[b]	
	Benchmark[c]	Narrative[c]	Strategy Map[c]	Strategy Map—Benchmark	Strategy Map—Narrative
1. The performance measures were usefully categorized in this case.	2.85***	2.93***	3.68***	0.83***	0.75**
2. The emphasis on financial measures was appropriate.	2.57***	3.42***	3.38***	0.82***	−0.03
3. The two business units, The Women's Store and The Family Store, used *some different* performance measures.	1.67***	2.23***	2.58***	0.92**	0.35
4. It was *appropriate* for the two business units, The Women's Store and The Family Store, to employ *the different* performance measures.	2.73***	3.05***	3.87***	1.13***	0.82***
5. The case was *easy to understand.*	2.37***	2.03***	2.45***	0.08	0.42
6. The case was *difficult to do.*	−1.48***	−1.30***	−1.02***	0.46	0.28
7. The case was *realistic.*	2.67***	3.30***	3.27***	0.60	0.03

** and *** indicate significance levels of 5% and 1%, respectively.
[a] Students were asked to rate their agreement with each statement on an 11-point scale anchored at 0.
[b] Means test of null hypothesis that mean for Strategy Map Treatment equals mean for Narrative Treatment or Benchmark Treatment.
[c] Test of whether mean response for participants in treatment equals 0.

Table 3
Analysis of the effects of treatment on participants' understanding of strategy.
To confirm our expectation that participants in the Strategy Map Treatment have a better understanding of strategy than those in other treatments, we asked participants to rate their agreement with four statements (1 through 4, below) designed to capture the essential elements of each SBU's strategy. Strategy Comprehension, the factor score from a factor analysis of Statements 1 through 4, provides a single, comprehensive measure of participants' understanding of strategy.

Statement[a]	Mean responses by treatment			Across-treatment differences in responses (*Z-score for non-parametric tests*)		
	Benchmark[b]	Narrative[b]	Strategy Map[b]	Strategy Map—Benchmark[c]	Strategy Map—Narrative[c]	Narrative—Benchmark[c]
1. The strategy of The Women's Store is to generate greater sales through its *existing* infrastructure rather than invest in new stores.	2.87***	2.93***	3.67***	0.80*** *3.67****	0.73** *3.31****	0.07 *0.44*
2. To grow sales, The Women's store must successfully introduce *new lines* of clothing to its *existing customers.*	2.55***	3.42***	3.37***	0.82*** *2.32****	−0.05 *−0.39*	0.87** *2.59****
3. The strategy of The Family Store is to grow by adding *new stores.*	2.67***	2.77***	3.27***	0.60 *2.64****	0.50 *2.01***	0.10 *0.59*
4. The Family Store needs an innovative marketing group because its growth plans depend upon the success of its advertising campaign in attracting *new* customers.	2.73***	3.05***	3.87***	1.13*** *3.44****	0.82*** *2.91****	0.32 *0.84*
Strategy Comprehension—factor score from factor analysis of Statements 1–4.	2.72***	3.03***	3.56***	0.84*** *4.01****	0.53*** *2.59****	0.31 *1.47*

"*" refer to significance levels for two-sample means test (top row) and Wilcoxon test that responses to Statements 1–4 and *Strategy Comprehension* for the Strategy Map Treatment exceed corresponding responses from narrative and Benchmark Treatments (one-tailed). Tests comparing responses from Narrative and Benchmark Treatments are two-tailed.
** and *** indicate significance levels of 5% and 1%, respectively.
[a] Students were asked to rate their agreement with each statement on an 11-point scale anchored at 0.
[b] "*" refer to significance levels of test that mean response (or factor) for specified treatment equals 0 (two-tailed).
[c] Top number in each pair is mean difference in responses across treatments. Lower number, in *italics*, is Z-score for Wilcoxon two-sample rank-sum test.

of The Women's Store. Only participants in the Strategy Map Treatment gave significantly higher performance ratings to the SBU that dominated on linked measures (Table 1). This suggests that only participants with strategy maps recognized the strategic significance of linked measures.

4. Results

4.1. Manipulation checks

After completing the performance evaluation exercise, we asked participants to rate, on an 11-point scale centered at zero, their agreement with seven statements assessing their understanding of the performance measures and the task (Table 2). Statements one through four address performance measure categorization and appropriateness. Manipulation checks show that participants believed that "measures were usefully categorized," that the "emphasis on financial measures was appropriate," that the two SBUs "used some different performance measures," and that use of different performance measures was "appropriate" ($p<0.01$). Agreement ratings for participants in the Strategy Map Treatment are significantly higher than ratings of participants in the Benchmark Treatment for these four statements ($p<0.05$, $p<0.01$). Agreement ratings for the Strategy Map Treatment significantly exceed ratings for the Narrative Treatment regarding the usefulness of performance measure categorization ($p<0.01$) (Statement 1) and the appropriateness of SBU-specific measures ($p<0.01$) (Statement 4), suggesting that participants who received strategy maps were better able to evaluate performance measure characteristics. Analysis of the last three statements indicates that all participants found the case easy to understand, not too difficult to complete, and realistic ($p<0.01$). Additionally, responses to the last three statements do not differ across treatments.

Table 4
Analysis of the effect of treatment and demographic factors on participants' understanding of strategy (*Strategy Comprehension*).[a,b]
We conduct the regression analysis modeled below to confirm our expectation that participants in the Strategy Map Treatment have a better understanding of strategy than those in other treatments and to determine whether demographic factors affect strategy comprehension.

$$\begin{aligned} \textit{Strategy Comprehension} = {} & \alpha_0 + \alpha_1\ \textit{Treatment B} + \alpha_2\ \textit{Treatment N} + \gamma_1\ \textit{Male}_i + \gamma_2\ \textit{Age} + \gamma_3\ \textit{WorkExp}-\textit{A\&A} \\ & + \gamma_4\ \textit{WorkExp}-\textit{M\&S} + \gamma_5\ \textit{Retail Experience} + \gamma_6 \textit{Masters}-\textit{M\%S} + \gamma_7 \textit{Masters}-\textit{M\%S} \\ & + \gamma_8 \textit{Years in US}>5 + \gamma_9\ \textit{Work Experience} + \gamma_{10} \textit{Number of Retail Visits} + \varepsilon, \end{aligned}$$

where the Strategy Map Treatment is the omitted treatment.

Variable	*Coefficient estimate*	*t-statistics*
Intercept	4.29	4.95***
Treatment B	−0.96	−4.28***
Treatment N	−0.67	−3.01***
Gender: Male	0.07	0.35
Age	−0.04	−1.32
Work Experience	0.04	1.47
Retail Experience	−0.43	−1.58
Years in US ≥ 5	0.47	2.15**
Number of Retail Visits	−0.00	−0.19
WorkExp-A&A	0.19	0.53
WorkExp-M&S	0.28	0.87
Major-A&A	0.15	0.44
Major-M&S	−0.25	−0.79
F-statistic	2.36***	
Adjusted R^2	8.34%	

** and *** indicate two-tailed significance levels of 5% and, 1%, respectively.

[a] *Strategy Comprehension* is a comprehensive measure of participants' understanding of each SBU's strategy. It is a factor score from a factor analysis of the following statements. 1) The strategy of The Women's Store is to generate greater sales through its existing infrastructure rather than invest in new stores. 2) To grow sales, The Women's Store must successfully introduce new lines of clothing to its existing customers. 3) The strategy of The Family Store is to grow by adding new stores. 4) The Family Store needs and innovative marketing group because its growth plans depend upon the success of its advertising campaign in attracting new customers.

[b] *Treatment B = 1* if participant was in the Benchmark Treatment and 0 otherwise. *Treatment N = 1* if participant was in the Narrative Treatment and 0 otherwise. *Male = 1* if the participant is male and 0 otherwise, *Age* = participant's age. *Work* Experience = the participant's full time work experience in number of years. *Retail Experience = 1* if the participant had work experience with a retail clothing store and 0 otherwise.*Years in US ≥ 5 = 1* if the participant was born in the US or had lived in the US for at least 5 years and 0 otherwise. *Number of Retail Visits* = the number of times the participant had visited a retail clothing store in the past 12 months. *WorkExp-A&A = 1* if participant's full time work experience is related to accounting, auditing or taxation, and 0 otherwise. *WorkExp-M&S = 1* if participant's full time work experience is related to marketing or sales, and 0 otherwise. *Major-A&A = 1* if participant's likely area of emphasis in the MBA program is accounting, auditing or taxation, and 0 otherwise. *Major-M&S = 1* if participant's likely area of emphasis in the MBA program is marketing or sales, and 0 otherwise.

4.2. Preliminary analysis

Our hypothesis is predicated on the assumption that strategy maps enhance strategy comprehension; therefore, we tested this assumption by asking participants to rate their agreement with four statements designed to capture the essential elements of each SBU's strategy.[7] Table 3 contains the strategy statements and the average agreement rating for each statement by treatment. Agreement ratings for the Strategy Map Treatment significantly ($p<0.01$ to $p<0.10$, one-tailed) exceed those of the Benchmark Treatment for all four questions. For three of the four questions, agreement ratings from the Strategy Map Treatment also significantly ($p<0.01$ to $p<0.10$, one-tailed) exceed those of the Narrative Treatment. We factor analyzed the four statements to develop a single measure of strategy comprehension, denoted *Strategy Comprehension*.[8] Consistent with our expectations, *Strategy Comprehension* for the Strategy Map

[7] The four strategy statements are: *1) the strategy of The Women's Store is to generate greater sales through its existing infrastructure rather than invest in new stores, 2) to grow sales, The Women's Store must successfully introduce new lines of clothing to its existing customers, 3) the strategy of The Family Store is to grow by adding new stores,* and *4) the Family Store needs an innovative marketing group because its growth plans depend upon the success of its advertising campaign in attracting new customers.*

[8] Responses loaded on a single common factor and Cronbach's coefficient alpha for the statements was 0.58, indicated internally consistency among the statements. An average of the four statements produced materially consistent results in subsequent testing.

Table 5
Comparisons of mean impact of linked and non-linked measures across information treatments.

BSC Combinations[b]		Mean of ratings differences (W−F) by favored (non-favored) measure type and treatment			H: Across-treatment comparison of impact of linked and non-linked measures (significance)[a]	
		(1) Benchmark Treatment	(2) Narrative Treatment	(3) Strategy Map Treatment	(4) = (3)−(1) Strategy Map—Benchmark	(5) = (3)−(2) Strategy Map—Narrative
(A) 1 and 2	Linked favor The Women's Store	0.833	1.000	1.167		
(B) 3 and 4	Linked do not favor The Women's Store[c]	−0.667	−0.633	−1.067		
(C) = (A)–(B)	*Mean impact* of linked measures[d]	1.500	1.633	2.233	0.733 (0.086)	0.600 (0.130)
(D) 1 and 3	Non-linked favor The Women's Store	0.633	0.467	0.267		
(E) 2 & 4	Non-linked do not favor The Women's Store[c]	−0.467	−0.100	−0.167		
(F) = (D)–(E)	*Mean impact* of non-linked measures[d]	1.100	0.567	0.433	−0.667 (0.164)	−0.133 (0.427)

** and *** indicate one-tailed significance levels of 5% and 1%, respectively.

[a] Across-treatment impact of linked (non-linked) measures in the Strategy Map Treatment compared with linked (non-linked) measures in the Benchmark and Narrative Treatments. Significance levels refer to two-sample, one-sided tests.

[b] BSC Combination 1: Link-WS, Non-WS; BSC Combination 2: Link-WS, Non-FS; BSC Combination 3: Link-FS, Non-WS; BSC Combination 4: Link-FS, Non-FS.

[c] Since the experiment is symmetric, "Linked do not favor The Women's Store" is equivalent to "Linked favor The Family Store" and "Non-linked do not favor The Women's Store" is equivalent to "Linked favor The Family Store."

[d] The *mean impact* is calculated by subtracting ratings differences for scorecard versions in which The Women's Store is favored on all linked (non-linked) measures from ratings differences for versions in which The Women's Store is not favored on any linked (non-linked) measures for each treatment.

Table 6
Across treatment comparisons of mean impact of linked and non-linked, controlling for participant demographic characteristics.
The regression model below is used to estimate the mean impact of linked and non-linked measures within each treatment, while controlling for demographic characteristics. Coefficient estimates are obtained by conducting regression analysis with 180 observations corresponding to the 180 participant responses. Using these coefficient estimates, the mean impact is calculated by taking the difference between the estimated value of the regression expression when the variable of interest (Link-WS or Non-WS) equals one and when it equals zero, calculated at the mean values of other variables.[a]

$$(W-F) = \alpha_0 + \alpha_1\ Link-WS + \alpha_2\ Non-WS + \alpha_3\ Treatment\ SM + \alpha_4\ Treatment\ N + \alpha_{12}\ Link-WS\tilde{n}Non-WS + \alpha_{13}\ Link-WS\tilde{n}Treatment\ SM + \alpha_{14}\ Link-WS\tilde{n}Treatment\ N + \alpha_{23}\ Non-WS\tilde{n}Treatment\ SM + \alpha_{24}\ Non-WS\tilde{n}Treatment\ N + \alpha_{123}\ Link-WS\tilde{n}Non-WS\tilde{n}Treatment\ SM + \alpha_{124}\ Link-WS\tilde{n}Non-WS\tilde{n}Treatment\ N + \gamma_1\ Male_i + \gamma_2\ Age + \gamma_3\ WorkExp-A\&A + \gamma_4\ WorkExp-M\&S + \gamma_5\ Retail\ Experience + \gamma_6 Masters-M\&S + \gamma_7 Masters-M\&S + \gamma_8 Years\ in\ US>5 + \gamma_9\ Work\ Experience + \gamma_{10} Number\ of\ Retail\ Visits + \varepsilon$$

Calculation of mean impact[b]	Estimated Mean Impact		Significance of Difference in Mean Impacts
	First Term[c]	Second Term[c]	
Impact L_{SM} > *Impact* L_B	2.307	1.489	0.043**
Impact L_{SM} > *Impact* L_N	2.307	1.484	0.041**
Impact N_{SM} < *Impact* N_B	0.503	1.136	0.089
Impact N_{SM} < *Impact* N_N	0.503	0.670	0.362

** and *** indicate one-tailed significance levels of 5% and 1%, respectively.

[a] Variable definitions: $W-F$ is the difference in evaluation scores between the Women's Store and The Family Store; *Link-WS* = 1 if performance on linked measures was higher for The Women's Store and 0 otherwise, *Non-WS* = 1 if performance on non-linked measures was higher for The Women's Store and 0 otherwise; *Treatment N* = 1 if a participant was a member of the Narrative Treatment, 0 otherwise; *Treatment SM* = 1 if the participant was a member of the Strategy Map Treatment and 0 otherwise; *Gender* = 1 if the participant is male and 0 otherwise, *Age* = the participant's age; *Work Experience* = the participant's full time work experience in number of years, *Retail Experience* = 1 if the participant had work experience with a retail clothing store, Years in US > 5 = 1 if the participant was born in the US or had lived in the US for at least 5 years and 0 otherwise; Number of Retail Visits = the number of times the participant had visited a retail clothing store in the past 12 months; WorkExp-A&A = 1 if participant's full time work experience is related to accounting, auditing or taxation, and 0 otherwise; WorkExp-M&S = 1 if participant's full time work experience is related to marketing or sales, and 0 otherwise; Major-A&A = 1 if participant's likely area of emphasis in the MBA program is accounting, auditing or taxation, and 0 otherwise; Major-M&S = 1 if participant's likely area of emphasis in the MBA program is marketing or sales, and 0 otherwise.

[b] *Impact* $L_{TREATMENT}$ and *Impact* $N_{TREATMENT}$ are the estimated mean impacts of linked and non-linked measures, respectively, in the *TREATMENT* indicated. For example, to calculate the estimated mean impact of linked measures in the Strategy Map Treatment, denoted *Impact* L_{SM}, we estimate the regression in Eq. (2) when the Women's Store dominates on linked measures in the Strategy Map Treatment (*Link-WS* = 1, *Treatment SM* = 1) and when the Women's Store does not dominate on linked measures in the Strategy Map Treatement (*Link-WS* = 0, *Treatment SM* = 1).

[c] "First Term" refers to the value of the first term in the expression in each row of the first column ("Calculation of Mean Impact") and "Second Term" refers to the value of the second term. For example, in the first row, the first term reports the estimated mean impact of linked measures in the Strategy Map Treatment, and the second term refers to the estimated mean impact of non-linked measures in the Benchmark Treatment.

Treatment (mean = 3.56) significantly exceeds ($p<0.01$, one-tailed) *Strategy Comprehension* for the Narrative Treatment (mean = 3.03). Thus, univariate tests are consistent with the expectation that participants who received strategy maps had a better understanding of strategy than those who did not.

Following Banker et al. (2004), we also compare strategy comprehension across treatments after controlling for demographic variables that may affect participants' understanding of strategy, such as age, gender, U.S. residency, and work experience. We perform this comparison by regressing *Strategy Comprehension* on indicator variables for the Benchmark and Narrative Treatments and a set of demographic control variables (Table 4). The omitted treatment is the Strategy Map Treatment, thus coefficients on the Benchmark and Narrative Treatment indicator variables (*Treatment B*, *Treatment SM*) estimate the effect that the information provided in these treatments has on strategy comprehension relative to information provided in the Strategy Map Treatment. The negative and significant ($p<0.01$, one-sided) coefficients on both of the treatment variables indicate that students in the Narrative and Benchmark Treatments did not understand strategy as well as those in the Strategy Map Treatment. An F-test of the control variables indicates that we can reject the hypothesis that none of the control variables differs significantly from zero ($p=0.02$); hence, inclusion of the control variables provides a more precise estimate of the treatment effects. Accordingly, we include control variables in tests of our hypothesis.

4.3. Hypothesis tests

We hypothesize that participants in the Strategy Map Treatment will place greater (less) weight on strategically linked (non-linked) measures in evaluating performance than those in the other treatments. Recall that the dependent variable is the difference in evaluation scores assigned to each manager by same participant, denoted $(W_i - F_i)$. To investigate the hypothesis, we compare ratings differences for participants who received the same BSC combination, but were in different information treatments.

The first set of tests provides simple, cross-sectional comparisons that are easy to interpret and have directional significance. In Table 5, we compare the mean impact of linked and non-linked measures across treatments. The mean impact of linked measures (Table 5, Row C) is calculated by subtracting the mean ratings differences for BSC combinations in which The Women's Store is not favored on any linked measures (Row B, BSC Combinations 3 and 4) from the mean ratings differences for combinations in which The Women's Store is favored on all linked measures (Row A, BSC Combinations 1 and 2).[9] As shown in Column 4 of Table 5, the mean impact of linked measures in the Strategy Map Treatment (2.233) marginally exceeds ($p = 0.086$, one-tailed) the impact of linked measures in the Benchmark Treatment (1.500). The mean impact of linked measures in the Strategy Map Treatment also exceeds ($p = 0.130$, one-tailed) that of linked measures in the Narrative Treatment (1.633), but this difference is not statistically significant. The mean impact of non-linked measures is contained in Row F. Consistent with the hypothesis, the mean impact of non-linked measures in the Strategy Map Treatment (0.433) is less than that of non-linked measures in the Narrative (0.567) and Benchmark (1.100) Treatments; however, the difference is not statistically significant.

We also use multivariate analysis to test our hypothesis because simple means tests do not measure interaction effects or control for factors outside of the single cross-section of interest. Repeated measures ANOVA allows for interactions among scorecard combinations and information treatments; however, ANOVA cannot accommodate continuous variables capturing important participant characteristics that preliminary tests have shown to be significant. Following Banker et al. (2004), we overcome this limitation by using regression analysis to compute the mean effects of linked and non-linked measures.[10]

We regress the dependent variable, $W_i - F_i$, on indicator variables that capture the four scorecard combinations, the three information treatments, all possible higher order interactions between measure type and treatment, and a set of control variables that capture participant-specific characteristics (Table 6, Eq. (2)). The regression model has moderate explanatory power (adjusted-$R^2 = 22.07$) and is statistically significant ($p<0.01$). An F-test of the model rejects the null hypothesis that all of the coefficients on the control variables equal zero ($p<0.05$). In particular, a control variable for gender is positively and significantly ($p<0.01$, two-tailed) associated with the ratings difference. Also, coefficients on control variables identifying participants with retail experience and those who have lived in the U.S. for at least five years are both statistically significant ($p<0.10$, two-tailed). Hence, including the control variables in the analysis yields more precise estimates of the impact of measure type and treatment on performance evaluation.

To test the hypothesis, we first compute the mean impact of linked and non-linked measures in each treatment using coefficients estimates and predicted values from the regression model in Eq. (2) of Table 6. Coefficient estimates are obtained by running the model with 180 observations obtained from the 180 experiment participants. Using these coefficient estimates, the mean impact is calculated by taking the difference between the value of the regression expression when the variable of interest (*Linked* or *Non-linked*) equals one and when it equals zero, calculated at the mean values of other variables (e.g., Govindarajan and Gupta, 1985).[11] For example, to calculate the estimated mean impact of linked measures in the Strategy Map Treatment, denoted *Impact* L_{SM}, we estimate the regression when the Women's Store dominates on linked measures in the Strategy Map Treatment (*Link-WS* = 1, *Treatment SM* = 1) and when the Women's Store does not dominate on linked measures in the Strategy Map Treatment (*Link-WS* = 0, *Treatment SM* = 1). The difference in these estimates is the mean impact of linked measures for

[9] Since the experiment is symmetric, The Family Store is favored on linked (non-linked) measures when The Women's Store is not favored on linked (non-linked) measures.

[10] ANOVA and multiple regression are based upon the same statistical model; therefore, results of the regression (without inclusion of demographic variables) are consistent with those of ANOVA (Keppel and Zedeck, 1989, Chapter 25). Without the demographic variables, the regression results are also identical to the means tests.

[11] Results for the regression displayed in Table 6 are available from the authors upon request. Only the mean effects of variables relevant to the hypothesis are included in Table 6.

participants with strategy maps. The mean impact of linked and non-linked measures is similarly calculated for the Benchmark and Narrative Treatments, and mean impacts are compared across treatments.

Consistent with our hypothesis, the estimated mean impact of linked measures in the Strategy Map Treatment (2.307) statistically significantly exceeds the corresponding mean impact in the Benchmark Treatment (1.489, $p = 0.043$, one-tailed) and the Narrative Treatment (1.484, $p = 0.041$, one-tailed). Also consistent with the hypothesis, the mean impact of non-linked measures is smaller in the Strategy Map Treatment (0.503) than it is in the Benchmark Treatment (1.136) and Narrative Treatment (0.670). However, only the difference between the Strategy Map and Benchmark Treatments is not statistically significant ($p = 0.084$, one-tailed). Together, the results indicate that participants receiving strategy maps appropriately placed significantly more weight on linked measures, but they did not place significantly less weight on non-linked measures.

5. Concluding remarks

We conducted an experiment to determine whether strategy maps improve a manager's ability to effectively use the BSC. Our results suggest that strategy maps enable participants to recognize the importance of strategically linked measures and appropriately place greater weight on these measures in evaluating performance. In particular, participants who received strategy maps placed more weight on measures linked to strategy than participants who received only narrative strategy descriptions and/or general company overview descriptions. This effect was more pronounced and significant after controlling for demographic characteristics and work experience.

Our results are subject to certain limitations. First, not all experiment participants had experience in the retail clothing industry, and none had extensive experience with the balanced scorecard. However, manipulation checks suggest that participants adequately understood the case materials and the task they were asked to perform. Second, our participants were not subject to the same incentives as managers normally responsible for evaluating subordinates' performance are. Third, while the experiment controlled performance on linked and non-linked measures, perceived performance relative to the targets could potentially differ across the measurement groups. Pilot tests and manipulation checks did not reveal any such problems; however, this possibility cannot be ruled out. Finally, the simplicity of the strategy map employed in the experiment potentially limits the ability to generalize these results to more complex maps. We use strip maps, which present a clear sequence of cause–effect relations between performance measures. When there is certainty about the relationships between the measures, strip maps avoid distractions and invite efficient behavior (Fiol and Huff, 1992, 272). More complex maps include information about the context in which a strategy is to be executed, such as competitive or regulatory factors. We chose a strip map because more complex maps may cause information overload and reduce decision makers' ability to interpret causal relationships in the data. Future research could investigate the decision-relevance of different strategy map features and designs.

Despite these limitations, this study should be of interest to the many managers who use balanced scorecards, as well as consultants who help implement them. In particular, the very feature that makes the BSC such a powerful tool for performance measurement, its strategic underpinnings, also makes it difficult to use. Temporally separated measures (*i.e.* leading and lagging measures) and non-linear causal relations inherent in non-financial measures impose significant cognitive demands on a user's working memory. Our results suggest that strategy maps can improve knowledge organization and thereby reduce the cognitive difficulty of using the BSC.

References

Adams D, Nelson R, Todd P. Perceived usefulness, ease of use and usage of information technology: a replication. MIS Q 1992;16:227–47.
American Institute of Certified Public Accountants, Maisel L. Performance measurement practices survey results. New York, NY: AICPA; 2001.
Ashton RH. Effects of justification and a mechanical aid on judgment performance. Organiz Behav Human Decis Processes 1992;52:292–306.
Axelrod R, editor. Structure of decision: the cognitive maps of political elites. Princeton: Princeton University Press; 1976.
Banker RD, Chang H, Pizzini MJ. The balanced scorecard: judgmental effects of performance measures linked to strategy. Account Rev 2004;79(1):1-23.
Barad M, Dror S. Strategy maps as improvement paths of enterprises. Int J Prod Res 2008;46(23):6627–47.

Benbasit I, Lim KH. The effect of multimedia on perceived equivocality and perceived usefulness of information systems. MIS Q 2000;24(3):449–71.
Brancato C. New performance measures—a research report. Report number 1118–95. New York, NY: The Conference Board; 1995.
Bryant L, Jones DA, Widener SK. Managing value creation within the firm: an examination of multiple performance measures. J Manage Account Res 2004;16:107–31.
Burney L, Widener SK. Strategic performance measurement systems, job-relevant information, and managerial behavioral responses-role stress and performance. Behav Res Account 2007;19:43–69.
Capelo C, Dias JF. A system dynamics-based simulation experiment for testing mental model and performance effects of using the balanced scorecard. Syst Dynamics Rev 2009;25(1):1-34.
Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13:319–40.
Davis FD, Yi M. Improving computer skill training: behavior modeling, symbolic mental rehearsal, and the role of knowledge structures. J Appl Psychol 2004;89:509–23.
Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989;35:982-1003.
Desanctis G, Jarvenpaa SL. Graphical representation of accounting data for financial forecasting: an experimental investigation. Account Organ Soc 1989;14(5–6):504–25.
Dilla WN, Steingbart PJ. The effects of supplementary display formats on balanced scorecard judgments. Int J Account Inf Syst 2005;6:159–76.
Druckenmiller DA, Acar W. An agent-based collaborative approach to graphing causal maps for situation formulation. J Assoc Inf Syst 2009;10(Special Issue):221–51.
Eccles R. The performance measurement manifesto. Harv Bus Rev 1991;69:131–7.
Eden C. On the nature of cognitive maps. J Manag Stud 1992;29:261–5.
Eden C, Ackerman F, Cropper S. The analysis of cause maps. J Manag Stud 1992;29(3):309–24.
Einhorn HJ, Hogarth RM. Decision making under uncertainty. J Bus 1986;59(4):225–51.
Eining MM, Jones DR, Loebbecke J. Reliance on decision aids: an examination of auditors' assessment of management fraud. Auditing: J Pract Theory 1997(Fall):1-19.
Epstein M, Manzoni JF. The balanced scorecard and tableau de bord: translating strategy into action. Manag Account 1997;79(2):28–36.
Fiol CM, Huff AS. Maps for managers: where are we? Where do we go from here? J Manag Stud 1992;29(3):267–85.
Fisher J. Use of nonfinancial performance measures. In: Young SM, editor. Readings in management accounting. Englewood Cliffs (NJ): Prentice Hall; 1995. p. 329–35.
Frownfelter-Lohrke C. The effects of differing information presentations of general purpose financial statements on users' decisions. J Inf Syst 1998;12(2):99-107.
Govindarajan V, Gupta A. Linking control systems to business unit strategy: impact on performance. Account Organiz Soc 1985;10:51–66.
Ittner CD, Larcker DF. Coming up short on nonfinancial performance measurement. Harv Bus Rev 2003:88–95. November.
Ittner CD, Larcker DF, Meyer M. Subjectivity and the weighting of performance measures: evidence from a balanced scorecard. Account Rev 2003a;78(3):725–58.
Ittner CD, Larcker DF, Randall T. Performance implications of strategic performance measurement in financial services firms. Account Organ Soc 2003b;28:715–41.
Johnson-Laird P. Mental models. Cambridge, U.K.: Cambridge University Press; 1983
Kaplan RS, Norton DP. The balanced scorecard—measures that drive performance. Harv Bus Rev 1992;70(1):71–9.
Kaplan RS, Norton DP. Putting the balanced scorecard to work. Harv Bus Rev 1993;74(5):134–47. 1993.
Kaplan RS, Norton DP. Using the balanced scorecard as a strategic management system. Harv Bus Rev 1996;74(1):75–85.
Kaplan RS, Norton DP. Having trouble with your strategy? Then map it. Harv Bus Rev 2000;78(5):167–76.
Kaplan RS, Norton DP. Transforming the balanced scorecard from performance measurement to strategic management: Part II. Account Horiz 2001;15(2):147–60.
Kaplan RS, Norton DP. Strategy maps: converting intangible assets into tangible outcomes. Boston, MA: Harvard Business School Press; 2004.
Kaplan RS, Norton DP. Alignment: using the balanced scorecard to create corporate synergies. Boston, MA: Harvard Business School Press; 2006.
Keppel G, Zedeck S. Data analysis for research designs. New York, NJ: W. H. Freeman and Company; 1989.
Larkin JH, Simon HA. Why a diagram is (sometimes) worth ten thousand words. Cogn Sci 1987;11:65–99.
Libby T, Salterio SE, Webb A. The Balanced Scorecard: the effects of assurance and process accountability on managerial judgment. Account Organ Soc 2004;49(4):1075–94.
Lipe MG, Salterio SE. The balanced scorecard: judgmental effects of common and unique performance measures. Account Rev 2000;75(3):283–98.
Lipe MG, Salterio SE. A note on the judgmental effects of common and unique performance measures. Account Organ Soc 2002;75(3):283–98.
Mair S. A balanced scorecard for a small software group. IEEE Software 2002;19(6):21–7.
Malina MA, Selto FH. Communicating and controlling strategy: an empirical study of the effectiveness of the balanced scorecard. J Manag Account Res 2001;13:47–90.
Montibeller G, Belton V. Causal maps and the evaluation of decision options—a review. J Oper Res Soc 2006;57:779–91.
Montibeller G, Belton V, Ackermann F, Ensslin L. Reasoning maps for decision aid: an integrated approach for problem-structuring and multi-criteria evaluation. J Oper Res Soc 2008;59:575–89.
Rucci AJ, Kirn SP, Quin RT. The employee–customer–profit chain at Sears. Harv Bus Rev 1998:82–97. January–February.
Scavarda AJ, Bouzdine-Chameeva T, Goldstein SM, Hays JM, Hill AV. A methodology for constructing collective causal maps. Decis Sci 2006;37(2):263–83.
Sterman JD. Business dynamics: systems thinking and modeling for a complex world. New York: Irwin McGraw-Hill; 2000.
Subramanian GH. A replication of perceived usefulness and perceived ease of use measurement. Decis Sci 1994;25(5,6):863–74.

Tan FB, Hunter MG. The repertory grid technique: a method for the study of cognition in information systems. MIS Q 2002;26(1): 39–57.
Todd P, Benbasat I. The use of information in decision making: an experimental investigation of the impact of computer-based decision aids. MIS Q 1992:373–93. September.
Tuttle BM, Kershaw R. Information presentation and judgment strategy from a cognitive fit perspective. J Info Syst 1998;12(1):1-17.
Umanath NS, Vessey I. Multiattribute data presentation and human judgment: a cognitive fit perspective. Decis Sci 1995;25(5,6): 795–824.
Urrutia I, Eriksen SD. Application of the balanced scorecard in Spanish private health-care management. Meas Bus Excell 2005;9(4): 16–26.
Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: Four longitudinal field studies. Manag Sci 2000;46(2):186–204.
Vera-Muñoz SC, Shackell M, Buehner M. Accountants' usage of causal business models in the presence of benchmark data: a note. Contemp Account Res 2007;24(3):1015–38.
Vessey I. Cognitive fit: a theory-based analysis of the graphs versus tables literature. Decis Sci 1991;22(2):214–41.
Veth G. Mapping the path to the future. DM Review 2006;16(5):32–3.
Waldmann MR, Holyoak KJ, Fratianne A. Causal models and the acquisition of category structure. J Exp Psychol 1995;124(2): 181–206.
Webb A. Managers' commitment to the goals contained in a strategic performance measurement system. Contemp Account Rev 2004;21(4):925–58.
Wyman BG, Randel JM. The relation of knowledge organization to performance of a complex cognitive task. Appl Cognit Psychol 1998;12:251–64.
Young S, O'Byrne S. EVA® and value-based management. New York, NY: McGraw-Hill; 2001.